

国立国語研究所学術情報リポジトリ

「日本語学習者による日本語作文と、その母語訳との対訳データベース」：設計と応用

著者	宇佐美 洋
雑誌名	言語データベース：さまざまな視点からの構築
ページ	12-19
発行年	2001-12-20
シリーズ	国立国語研究所研究発表会；平成13年度
URL	http://doi.org/10.15084/00002939

「日本語学習者による日本語作文と、その母語訳との対訳データベース」 —設計と応用—

宇佐美 洋（日本語教育部門第一領域）

smudr@kokken.go.jp

1. データベースの概要

- 1) 日本語学習者が書いた日本語作文（400 字～800 字程度）
- 2) 1) を執筆者自身が自分の母語（あるいは最も楽に文章が書ける言語）に翻訳したもの
- 3) 1) の一部に、日本語教師（日本語を母語とする教師と、学習者と同じ母語をもつ教師の両方を含む）が添削をほどこしたもの
- 4) 執筆者・添削者の言語歴情報

以上 4 種類のデータを大量に集め、電子化した上で、簡便な検索が可能になるように html によるインデックスを付したもの。2001 年 3 月に CD-ROM の形式で正式公開。

データ収集国ごとに検索用のインデックスを html で作成。インデックスから各ファイルにリンクがはってあり、クリックによって任意のファイルを呼び出すことができる。

html によるインデックス例（韓国語母語話者データ検索用）

Index: Korea												
Back to the top page												
ID No.	Translated into:	JP-txt	JP-pdf	JP-jpg	MT-txt	MT-pdf	Writer's Info	Title of Essay	Revision by NS	Reviser's Info (NS)	Revision by NNS	Reviser's Info (NNS)
KR001	Korean	kr001i.txt	kr001i.pdf	kr001i.jpg kr001i2.jpg	*	*	kr001n.txt	2	*	*	*	*
KR002	Korean	kr002i.txt	kr002i.pdf	kr002i.jpg kr002i2.jpg kr002i3.jpg	kr002m.txt	kr002m.pdf	kr002n.txt	1	*	*	*	*
KR003	Korean	kr003i.txt	kr003i.pdf	kr003i.jpg kr003i2.jpg	kr003m.txt	kr003m.pdf	kr003n.txt	2	*	*	*	*
KR004	Korean	kr004i.txt	kr004i.pdf	kr004i.jpg kr004i2.jpg	kr004m.txt	kr004m.pdf	kr004n.txt	2	*	*	*	*
KR005	Korean	kr005i.txt	kr005i.pdf	kr005i.jpg kr005i2.jpg	kr005m.txt	kr005m.pdf	kr005n.txt	2	*	*	*	*
KR006	Korean	kr006i.txt	kr006i.pdf	kr006i.jpg kr006i2.jpg	kr006m.txt	kr006m.pdf	kr006n.txt	2	*	*	*	*
KR007	Korean	kr007i.txt	kr007i.pdf	kr007i.jpg kr007i2.jpg	kr007m.txt	kr007m.pdf	kr007n.txt	2	kr007ns1.jpg kr007ns2.jpg	p20n.txt	*	*
KR008	Korean	kr008i.txt	kr008i.pdf	kr008i.jpg kr008i2.jpg	kr008m.txt	kr008m.pdf	kr008n.txt	2	kr008ns1.jpg kr008ns2.jpg	p02n.txt	kr008nn1.jpg kr008nn2.jpg	kr01n.txt
KR009	Korean	kr009i.txt	kr009i.pdf	kr009i.jpg kr009i2.jpg	kr009m.txt	kr009m.pdf	kr009n.txt	2	kr009ns1.jpg kr009ns2.jpg	p31n.txt	*	*
KR010	Korean	kr010i.txt	kr010i.pdf	kr010i.jpg kr010i2.jpg	kr010m.txt	kr010m.pdf	kr010n.txt	2	*	*	*	*
KR011	Korean	kr011i.txt	kr011i.pdf	kr011i.jpg kr011i2.jpg	kr011m.txt	kr011m.pdf	kr011n.txt	2	*	*	*	*

2. データ収集国・データ数

日本を含むアジア 10 か国（インド・ヴェトナム・韓国・カンボジア・シンガポール・タイ・中国・マレーシア・モンゴル・日本）の日本語教育機関に協力を依頼し、日本語学習者の作文のべ 1,000 名分を収集。また比較のため、日本語母語話者の同課題による作文も収録。

3. データ形式

データ使用者の便を考え、データによっては同一種類のものを複数のファイル形式で保存したものもある。

- 1) 日本語作文（テキストファイル／pdf／画像(jpg)ファイル）
- 2) 母語訳（テキストファイル／pdf）
- 3) 執筆者情報（テキストファイル）
- 4) 日本語作文添削（jpg ファイル）
- 5) 添削者情報（テキストファイル）

添削情報は、現バージョンでは手書きのものをスキャナで画像として入力したものが収録されているが、現在 XML をもちいて添削情報も検索可能な形で電子化する方法の開発も進行中(7.2 節参照)。

4. 作文課題

課題 1: できごとを時間順に叙述する作文 「あなたの国の行事について」

課題 2: 自分の意見を論理的に述べる作文 「たばこについてのあなたの意見」

以上 2 課題から、学生自身、または教師が課題を選択

5. データベースの効能

- 1) 日本語学習者による作文が大量に収録されていることで、誤用分析・中間言語研究のための基本データベースとして使用可能
- 2) 日本語と母語等が対訳の形で示されていることで、対照言語学的研究に使用可能
- 3) 日本語作文とその母語等への訳を比較することで、母語干渉研究が可能
- 4) 短文でなく、まとまった分量の文章が収録されていることで、論理展開・文章構成に関する研究が可能
- 5) 多数の日本語教師による添削が収録されていることで、日本語学習者に対する作文添削の方法論についての研究が可能
- 6) 作文の評価や効果的的作文指導法に関する研究が可能

など。

6. データベース応用例

タイ語母語話者による日本語作文例

- (1) しかし、私はたばこを吸うの初めはやめられないまで、まやくをも
 っと強くするようになってとうとう死んでしまったと思います。
(th053¹)

本人による母語（タイ語）訳と、その逐語訳

- (2) tee chan khit waa kaan rəəm suup burii
 しかし 私 思う ～と こと 始める 吸う たばこ
- ca tham hay seep saan seep tit thii runreeng maak khuun
(未来) させる 中毒する 強く 多い さらに
- con may saamaat ləək day ləew
～まで (否定) できる やめる できる (完了)
- nay thii sut ko sia chiiwit²
 結局 失う 生命

タイ語から日本語への再翻訳

- (3) しかし、いったんたばこを吸い始めると中毒はだんだんひどくなり、ついには
 はやめられなくなって結局命を失うことになると思います。

タイ語“con”は、日本語「まで」と同様、動作・状態の到達点をあらわす。日本語教育でもタイ語教育でも、少なくとも初歩の段階では“con”＝「まで」と教えられることが多いが、(2)の状況を「まで」を使って表現しようとするとうまくいかない。

- (4) ??しかし、たばこを吸い始めることは、やめることができないまでだんだん
 ひどく中毒させ、結局命を失うことになると思います。

¹ 本データベースで用いられている執筆者 ID 番号である。最初の 2 文字がデータ収集国、その後の 3 文字が収集国内での執筆者番号を示す。以下同じ。

² 声調表示は省略。

(4)は日本語として「誤り」とはいえない。しかしこの文脈では、(3)のように「ついには」「とうとう」のような副詞的表現を用いた方がよい。→なぜ「まで」ではやや不自然？

日本語「まで」のもつ含意

「到達点」が一種の「完成点・最終点」としてあらかじめ想定されており、その完成点に達した時点において行為や状態が「少なくともいったんは終了」する

(5) 彼は酔っぱらうまで飲んだ

(5)には「酔うことを目的として飲んだ」、または「酔っぱらった時点で飲むのをやめた」という含意が感じられる。

(6) どうしてこんなにひどくなるまで放っておいたの？

(6)の場合、この場合、このまま放置をしつづければ今後もさらにひどくなる可能性があり、到達点以降で行為や状態が完全に終了するとは言えない。しかし話し手自身は「現在よりひどくなるというようなことはあってはならない」と考えており、その意味で現状を一種の「最終状態」と想定していることになる。

喫煙とたばこ中毒との関係はどうか…たばこ中毒は、たばこがやめられなくなって時点で終了するわけではない。その後もさらに進行する。

→この状況と「まで」のもつ含意とは合致しない。したがってやや不自然になる。

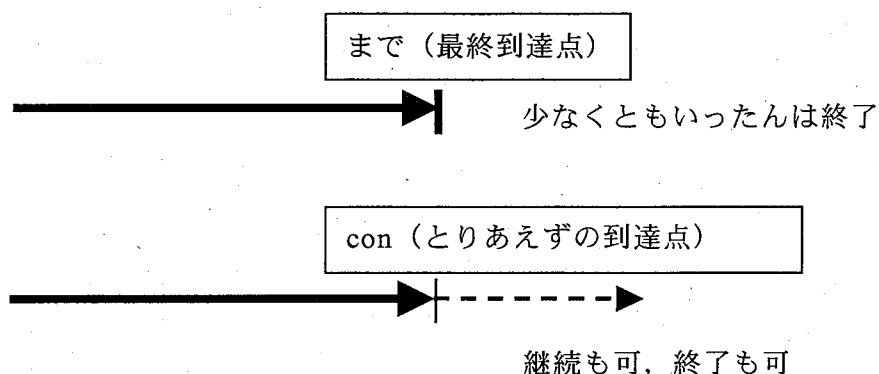
タイ語“con”，(2)の文脈でも使用可能…タイ語の“con³”には、「完成点・最終点」という含意はない。

(5') khaw duum con maw
かれ 飲む まで 酔う

³ Kessakul and Methapisit(2000)は con を transition marker と呼び、以下のような定義を与えている。 “a syntactic entity used to denote a certain amount of time needed for an action in V₁ to progress before going into the endpoint of state signified by a result verb in V₂”. つまり、第1動詞であらわされる行為が、第2動詞のあらわす状態へと「ある時間をかけて推移する」というところにこの語の本質があるという考え方である。

(5')では、酔っぱらった時点で飲むのをやめることもありうるし、やめないこともありうる。

日本語「まで」とタイ語“con”の違い



→タイ語母語話者の書いた不自然表現をもとに、なぜそのような表現が生まれたのか、日本語としてなぜそれが不自然なのかを考察することで、日本語・タイ語それぞれについてより深く分析するための手がかりとなる。

7. 今後の発展

7.1 欧米言語を対象とする対訳データベースの作成

本年度からデータ収集開始。2003年度試作版公開の予定

英語・フランス語・ドイツ語・スペイン語・ポルトガル語等のデータを収集の予定。

7.2 添削情報電子化システムの開発

「削除」「挿入」「置き換え」などの添削情報を、XMLのタグによってテキスト中に埋め込む。それぞれのタグで挟まれた部分の表示方法を「スタイルファイル」によって定義しておけば、タグの付与されたテキストをブラウザで読み込むことで、コンピュータの画面上でも手書きでの添削結果に似た形で表示させることができる。

タグ挿入例：

私は走って学校へ行きましたです。

意味：

タグとで挟まれた部分を削除

ブラウザ上での表示：

私は走って学校へ行きました ぞす。

「置き換え」をあらわすタグに3種類のを設定し、添削における添削者の「態度」を表示できるようにする。

1) rep(replace)タグ

書式：`<rep value="置き換え後の文字列">置き換え前の文字列</rep>`

「日本語として明らかな誤用であるため、直すべきだ」ということを表示するために使用

2) better タグ

書式：`<better value="より良い表現">元々の表現</better>`

「日本語として誤りとはいえないが、より適切な表現が存在する」ことを表示するために使用

3) doubt タグ

書式：`<doubt value="暫定的訂正表現">執筆意図不明の表現</doubt>`

「執筆意図が必ずしもよくわからないが、こういうふう直せばいいのではないかと、疑いながら添削をおこなうときに使用

タグに対して検索をかけることで、添削者の「添削態度」（ある表現を明らかな誤用と認めるか、そうでないか、など）について調査することが可能になる。

→より効果的な作文添削方法に関する調査研究へ。

「誤用」とは何か？

学習者の発達段階に応じて、何をどこまで訂正するのが効果的か？

日本語教師としての経験年数によって、添削のやり方にどのような変化が生じるか？

現在、「日本語学習者による日本語作文と、その母語訳との対訳データベース ver.2」に収録されている添削付作文は、すべてこの方式によるタグ付け作業完了。

7.3 音声による対照言語データベースの開発

現在作成準備中。

概要：

日本語学習者にあるタスクを与え、それに基づいてまず日本語で発話をしてもらい、その後同じタスクに基づいて、母語でも発話をしてもらう。発話を録音・文字化し、音声データとテキストデータとの間にリンクをはる。

収集予定のデータ：

	自由度低	自由度中	自由度高
モノ ロー グ	A. 朗読 (音声面での対照研究 のための基礎資料とし て)		C. テーマに基づく自由 な意見陳述 (意見陳述における各 種ストラテジー・論理構 成例の収集)
ダイ アロー グ		B. ひとりの発話内容を あらかじめ固定したダ イアログ (文法・語彙使用など における用例の収集)	D. インフォーマント2人 による、ロールカードに 基づくロールプレイ (ある場面におけるコ ミュニケーション行動 例の収集)

宇佐美 (2001) p.43

目的・効能：

- 1) 日本語と学習者の母語 X 語について、音声面での基礎データの収集が可能
- 2) 日本語学習者による日本語発話における、話し言葉での誤用・不自然表現の実例の収集が可能
- 3) 「作文対訳データベース」との対照により、音声面での中間言語と文章面での中間言語との対照が可能
- 4) 「音声による伝達」に関するさまざまなストラテジーの実例収集が可能
- 5) 特にロールプレイでは、学習者に「日本人役」を演じてもらうことで、「日本人のコミュニケーションストラテジーが、学習者にはどのようなものとして受け取られているか」ということについて知る手がかりが得られる。

今後の作業予定：

02-03 年度 データベース構築のための予備研究
04-05 年度 データベース試作版作成

☆参考文献

Kessakul, Ruetaivan and Tasanee Methapisit(2000) Resultative construction in Thai and the related issues. *Proceedings of the 5th International Symposium Languages and Linguistics: Pan-Asiatic Linguistics*, 16-17 November 2000

国立国語研究所(2001)『日本語学習者のための日本語作文と、その母語訳との対訳データベース ver.2』

宇佐美 洋(2001)「これからのスピーチ研究—日本語教育の立場から—」,『日本語学』vol.20, pp.37-47, 特集：スピーチの研究とその教育

宇佐美 洋(in print)「『対訳データベース』と日本語教育」,『日本語と外国語との対照研究シリーズ X 対照研究と日本語教育』, 国立国語研究所編, くろしお出版

「日本語学習者のための日本語作文と、その母語訳との対訳データベース ver.2」については、以下の URL からデータの一部をご覧になれます。

<http://www2.kokken.go.jp/~smudr/public/sakubun/index.htm>

またこのデータベースを収めた CD-ROM は、研究・教育目的のために使っていただける方には無料で配布しております。CD-ROM の申し込み方法についても、上記 URL に記載してあります。

さらに、このデータベースの試用版(ver.1, 2000 年 8 月公開)に基づく諸研究を収めた報告書『日本語教育のためのアジア諸言語の対訳作文データの収集とコーパスの構築』(平成 11-12 年度科学研究費補助金基盤研究(B)(2)研究成果報告書)も、上記 URL からダウンロードできるようになっています。ただしこれらの論文を読むには、アドビ社の無料ソフトウェア「アクロバットリーダー」ver.4 以上が必要です。